DOCUMENT RESUME

ED 215 250 CG 015 853

AUTHOR Janz, Tom
TITLE Towards a Performance Data and Development System: Getting Rid of Performance Appraisal.
PUB DATE Jun 81
NOTE 38p.; Paper presented at the Annual Conference of the Canadian Psychological Association (42nd, Toronto, Ontario, Canada, June 3-5, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Behavioral Science Research; *Behavior Rating Scales; Employer Employee Relationship; *Evaluation Criteria; *Evaluation Methods; Foreign Countries; Observation; *Organizational Development; Organizational Effectiveness; *Performance; *Personnel Evaluation; State of the Art Reviews
IDENTIFIERS *Canada

ABSTRACT

If organizations are to measure and use worker performance information effectively, they must distinguish between two components of performance appraisal: performance data (recorded information for comparing workers) and performance development (the process of improving human assets by discouraging ineffective and reinforcing effective job behaviors). Of the commonly used performance data types, unstructured and structured narratives suffer from a lack of common measures to compare workers as well as a high degree of rater subjectivity. Trait-defined scales offer structured measures but often have ambiguous dimensions and obscure methods for rating workers. Behaviorally anchored scales, developed to provide behaviorally specific definitions of dimensions and anchors, offer little advantage over trait-defined scales at considerable cost. End result data offer quantitative measures that reflect a worker's organizational contribution. If personnel development is a priority, however, a behavior observation scale (BOS) system may work best. A data and development system built around a BOS system can: (1) separate out the conflicting goals of clearly differentiating workers and insuring worker commitment to performance improvement; (2) create a data subsystem with steps to resolve conflicting views of performance levels; and (3) provide a development system that avoids direct comparisons and focuses on mutually acceptable, specific goal setting. (NRB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED215250

TOWARDS A PERFORMANCE DATA AND DEVELOPMENT SYSTEM: GETTING RID OF PERFORMANCE APPRAISAL

Tom Janz

School of Business Administration

Simon Fraser University

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Tom Janz

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CG 015853

Running Head: Performance Data and Development

TOWARDS A PERFORMANCE DATA AND DEVELOPMENT SYSTEM: GETTING RID OF PERFORMANCE APPRAISAL

Everyone knows that there is something wrong with perfcrmance appraisal. Evaluating workers makes supervisors nerveous and mean. This is because seeing their ratings causes workers to become angry and vindictive. Worse yet, they raise embarrasing questions about why they were given a particular rating. Often, neither the evaluators nor the evaluated feel like anything useful is accomplished, but everbody keeps doing it anyway, once a year. Workers are particularly bitter when ratings on vague, "motherhood" traits are used to justify higher bonuses for some. Where there is an absence of clearly defined performance measurement, supervisory favoritism or ingratiation are seen to lie behind the numbers. Workers are suspicious and distrusting of what resides in personnel files.

All of these feelings are symptoms of the problems facing current practice in performance data and appraisal, but they are symptoms and not the problems themselves. While many are aware of the symptoms, few have carefully examined what the root causes are and even fewer have outlined specific solutions to these prcblems.

This paper: a) separates out performance data and

performance development, 2) compares and evaluates basic types of performance data, 3) integrates the objectives of perfcrmance data and development, and 4) describes a behaviorally based system that maintains accurate performance data for personnel decision making and motivates maximum performance development.

Sorting out Performance Data from Personnel Development

While appraisal is typically treated as a unified whole, it is useful to separate the recording of performance data from the act of personnel development. Performance data is recorded information used to line up workers in terms of their individual effectiveness. Performance data assists managers in making personnel decisions. To do this effectively, the data must accurately and clearly reflect each workers individual contribution to the organization. Further, the data recording process must be seen to be fair. Having good data that is perceived to be biased can result in resentment that undermines and even wipes out any benefit from good decisions reached on the basis of that data. Performance development assists managers in the process of building human resource strength. To do this effectively, both workers and managers need a clear understanding of what constitutes effective and ineffective job behavior, and what specific behaviors are most in need of attention by a specific worker. Here, worker acceptance of the development

message and process is not just a bonus, it is essential. The kind of performance data maintained by an organization will have direct impact on the quality of its personnel decisions, and a secondary impact on its development procedures. Also, the kind of performance data gathered will impact on the acceptability of decisions made on the basis of that data. This is especially true for compensation and merit pay decisions. The next session compares six performance data types in terms of use, costs, worker acceptance and utility for assisting personnel decision making and guiding personnel development.

Performance Data Types: A comparison

Table 1 lays out a comparison of six performance data types. The following paragraphs expand and explain the table.

Insert Table 1 about here

"Blank page" performance data is nothing more than a subjective written evaluation of a worker's performance. It is most widely used for managerial level jobs. If measurement is defined as attaching numbers to entities for a purpose, blank page performance data is not measurement. Since it generates no numbers, it cannot be used to make personnel decisions. During performance appraisal, having a

Table 1 : A Comparison of Performance Data Types

Data Type	Est Use	Development Time	Development Reqs	Process Clarity	Acceptance	Utility for Assisting Personnel Decisions	Utility for Assisting Personnel Development
Unstructured Narrative "blank page"	15	0	1 man.	Topics covered and language used completely up to manager	All or none little opportunity for mutual bargining	none: no common ground for comparing workers	slight: memory aid during performance discussion.
Structured Narrative "report card"	30	2 hrs	1 man.	Topics guided by vague phrases:content up to manager	Some chance for bargining on individual topics	slight: workers compared on broad topics	some: written comments covering common ground could be used to build specific targets for improvement but criticism can lead to rejection
Trait Defined Adjective Anchored Scales	35	2 hrs	1 man.	Topics guided by vague phrases: numbers attached using ambiguous anchors	Little solid basis for justify -ing a rating	some:good numbers but poor acceptance	slight: numbers based on so mushy a procedure are a poor basis for obtaining worker commitment to behavioral improvement.
End Result Measures "scorecard"	15	2 hrs	1 man.	Numbers attached using counting operations. Some -times the process obscure (balance sheets)	High if end result measure relevant to worker input	good: IF contamination by other factors avoided	some: relevant end result measures point up the goal but say nothing about what to do to get there.
Behaviorally Anchored Scales	5	80 hrs	5 man.	Dimensions and anchors composed of specific behaviors but rating process unclear.	Hard to justify placing a worker on a BAS.	some: good numbers but still comes to be seen as subjective	slight: identifies specific effective and ineffective behavior but does not point up which behaviors each worker needs to improve most. Also, criticism leads to rej.
Behavior Observation Scales	0	80 hrs	1 man.	specific dimensions and anchors and clear rating process	High if opportunity given for mutual discussion behavioral frequencies	good: good numbers and the potential for hi acceptance	some: specific effective and ineffective behavior patterns identified but resistence to direct criticism still a problem.

written evaluation in front of the appraisor could help make sure that at least the performance topics contained in the data are covered in the appraisal session. Left with nothing in writing at all, appraisors could well forget to discuss contentious performance topics, preferring to avoid potential confrontations. Still, there are no guarantees that the appraisor has recorded key performance topics in the written evaluation and there is no built in system for communicating these observations to workers. Worker acceptance depends largely on the content of the report and the regard the worker has for his supervisor. The content of the report is not often up for discussion and it is hard to imagine how worker input could be worked in to this data type.

One rung up the ladder, the "report card" approach adds structuring to the data recording process. The rater still writes narrative comments about the rate, but the comments are guided by phrases such as greatest strength, work attitude, and overall assessement. While this increased structure represents some improvement over blank page data, it is still not quantitative and offers little direct help for making personnel decisions. There is no guarantee that the general phrases used to head up the report card strike at the key job dimensions. Further, these vague phrases offer wide room for differing interpretation by different

raters. Thus the report card approach offers little for personnel decisions and only somewhat more for performance appraisal. On the plus side, the raters recorded observations offer a rich basis for conducting a meaningful performance review discussion. Compared to a set of numbers or a series of checks on some vaguely anchored scales, there is much more to stimulate discussion in a set of structured, written comments. The risk is that the comments are irrelevant or wrong, but at least there is a basis of discussion. Numbers are so unrealistically final.

Trait defined, adjective-anchored scales are the next type of performance data in Table 1. While they produce quantitative data needed for personnel decision making, they represent a giant leap backward in the recording of performance data for personnel development. Performance topics are defined by vague, ambiguous phrases such as initiative, work quality, loyalty, and attitude. They do produce numbers, but what do the numbers mean? When the anchors used to attach numbers include such telling phrases as "meets expectations" or "somewhat above average", there is no concrete meaning that can be attached to a given worker's score on a particular dimension. Raters filling out such forms often are unwilling to rate workers very low or high, tending to obscure their strengths and weaknesses in a set of "average" scores. Still, when a worker's score

is averaged over several dimensions, these simple minded scales do as well as more carefully derived measures in separating effective from ineffective workers. So for making personnel decisions where worker acceptance is unimportant, the average dimension score over a set of trait-defined, adjective-anchored scales is useful. However, when worker acceptance of the personnel decisions has important organizational impacts, the drawbacks can easily outweigh the advantages. For example, if this relatively simple minded performance data is used to make compensation or merit decisions, resentment can surface in motivational behaviors such as willingness to accept overtime, initiating vs. avoiding customer service, or calms vs. further angers frustrated customers. For performance development, the numbers assigned to individual workers on a given scale can be worse than useless. The nature of what is being rated and the process of attaching numbers are both obscure.

End result measures are popular among engineers and accountants because they are hard numbers. It does not occur often enough to such managers that they may be the wrong numbers, or if not wrong, at least contaminated. Sales volume, met production quotas and under budget projects are common end result data. The problem is that a high score of these measures can often reflect factors other

than the efforts and skills of the manager or salesman involved. Differences in sales territories, production technology, or project complexity can obscure or even reverse the effects of a skillful, hard working manager or worker. Before using the hard numbers of the scorecard approach, a hard look at contaminating influences is needed, but is often missing. The numbers must directly reflect individual effort and skill, not the luck of the draw or irrelevant factors. For personnel decisions, where end result measures closely reflect an individual's contribution, they are ideal. Further, worker acceptance of such measures is often high although obvious sources of contamination cut into credibility. By contrast, for performance appraisal, scorecard data offers few guidelines to specific behaviors that need improvement. End result data clarifies the goal of improved performance, but says little about how to achieve that goal. For example, it may be of some help to know that a reasonable sales target is $500,000 gross, but that alone provides no message on what to do to reach that goal. A good performance data system for appraisal purposes must clarify the path as well as the goal.

So far, the traditional, cheap approaches to recording performance data have been reviewed. Blank page and report card formats offer performance detail and moderate worker

acceptance but suffer from a lack of common measures to compare those rated and a high degree of subjectivity from rater to rater. Trait-defined adjective-anchored scales offer structured measurement, but at the cost of ambiguous dimensions and obscure methods for attaching numbers to workers on those dimensions. While a worker's score averaged over several dimensions is useful for making personnel decisions, the same worker's score on each dimension is a poor indicator of his strengths and weaknesses. End result performance data offers quantitative measurement on clear dimensions but is often contaminated by factors other than an individual workers contribution to organizational effectiveness.

These traditional methods account for most of what is done in recording performance data. A quick look at the cost and personnel requirements points to one reason for resistence to a wholesale move into the behavioral methods. Certainly the behavioral methods have a heavy burden of proof to justify their increased costs. Still, personnel and top managers have been all too willing to overlook the costs of doing a poor job of recording performance data. The result is the sorry state of affairs pointed out at the start of this paper.

Behaviorally Anchored Scales (BAS) were developed to provide behaviorally specific definitions of dimensions and

anchors (Smith and Kendall, 1965). Figure 1 in the following section contains an illustration of a BAS for measuring the dimension "Working steadily vs. wasting time." The BAS process involves several time consuming steps: 1) gathering a pool of critical incidents -- stories about specific times that are examples of effective, ineffective, and average performance, 2) sorting those incidents into job dimensions, 3) re-sorting the incidents into the job dimensions defined in 2, 4) rating the effectiveness of the behaviors that define each dimension, 5) choosing a small number of incidents for rating anchors on the basis of how well raters agree that an incident belongs on a particular dimensions and how well they agree that the incident belongs at a particular level of effectiveness within that dimension. For all this effort, research comparing BAS with adjective-anchored scales finds small if any improvements in the characteristics of the ratings (Borman and Dunnette, 1975; Bernardin, 1979). Thus for personnel decisions, BAS offer little advantage at considerable cost. As input to appraisal, BAS often confuse the worker being rated. As is explained in more detail below, a worker can be rated as falling near a particular anchor without that worker ever performing the behavior contained in the anchor. Specific effective and ineffective behaviors are described, but the specific behavior most in need of improvement for a given

worker can easily be missed.

BAS were developed with worker and supervisor acceptance in mind, but something has clearly not worked out. Supervisors and workers have direct input in contributing incidents, and supervisors provide much of the data for forming dimensions and anchoring performance on those dimensions. Still, in the end, supervisors are asked to put an X on a behaviorally defined scale, anchored by specific incidents. The X placing task remains highly subjective and hard to verbally justify. A particular worker could have just performed one of the incidents that anchors poor performance, but that worker may not typically behave that way. Where should his X fall. Similarly, the place on the scale that best represents a particular worker's dimension position may be anchored by a behavior that worker performs rarely. This author's experience with BAS is that they are developed with gusto and are practised with resentment. Actually having to make several sets of ratings becomes more and more frustrating as the conflicting effects of recency and typicality of behavior are felt. Workers also experience frustration as supervisors are unable to verbally clarify the rating process or their specific ratings.

Behavior Observation Scales (BOS) were developed to provide a clearer behavioral picture of the appraisee's

problems and accomplishments than is possible with either adjective or behaviorally anchored scales. This behavioral picture can serve as a firm basis for performance data or development discussions where the pictures given by adjective or BAS are seen as lacking. Figure 1 diagrams the feedback that a detail design technician working

---

Insert Figure 1 about here

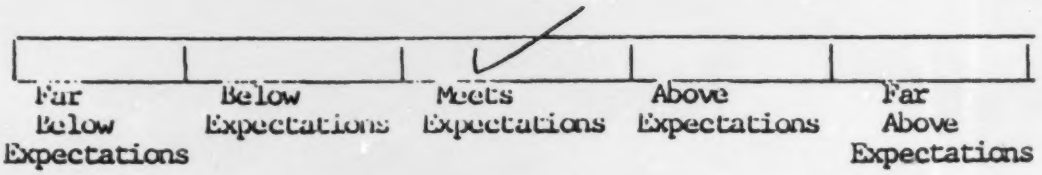
---

in a department of public works received from two rating systems on the same performance dimension. The top scale was the standard type used by the county. The BAS was developed by the author. Both scales placed the ratee near the middle of the scale on this dimension and near the middle of his peers as well. Certainly the BAS is more carefully defined and anchored, but does it point up what the supervisor should discuss with this ratee any better than the old standard? Maybe a little, but not much. In both cases, the information contained in the feedback suggests that this dimension need not enter a goal setting discussion, since the appraisee scored about average. At best, BAS the feedback suggests the supervisor bargain for higher targets with the effective behaviors and for reductions of the ineffective behaviors. Unfortunately, during the interview the supervisor is left to rely on his

TRAIT DEFINED- ADJECTIVE ANCHORED

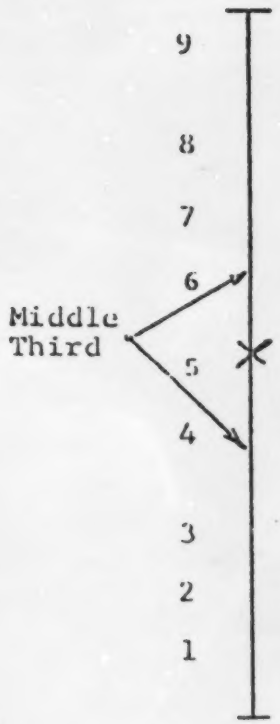
LIKERT TYPE

4. DILIGENCE



BEHAVIORALLY ANCHORED

Working Steadily vs. Wasting Time



- This technician could be expected to work steadily through his assignments until completed and then ask for new work.
- When finished his assignments, this technician could be expected to find some useful small job to do to fill in before the next assignment came in.
- This technician could be expected to complete assignments on schedule but seldom further work until it was assigned.
- This technician could be expected to spend about 30 minutes a day looking out the window or watching the clock.
- This technician could be expected to engage in long personal phone calls when there was work to be done.
- This technician could be expected to go around the office interrupting others for discussions on hockey, football, baseball, fishing, etc.

impressions and recollections concerning how often this technician acted in specific effective and ineffective ways. A more careful analysis outlined in the following section revealed that the technician was rated average NOT because his behavior was similar to anchor 5. Rather, his behavior was judged similar to the top anchors and just the opposite of most of the bottom anchors, but he made more long personal phone calls than any of the other technicians. The end result was a middle of the scale rating for both standard and BAS that obscured instead of pointed up the specific behavioral problem.

In short, neither the county scale nor BAS provide an adequate framework for structuring an effective performance data or development discussion. The difference between BAS and BOS is that instead of rating a worker on a dimension, each worker is assessed against each of the principle kinds of effective and ineffective job behaviors. Keeping in mind the dimersion diagramed in Figure 1, Figure 2 contains the results of a BOS run for the same worker on the same dimension.

Insert Figure 2 about here

The top section of Figure 2 lists the five behavior paragraphs that define the dimension. The second part

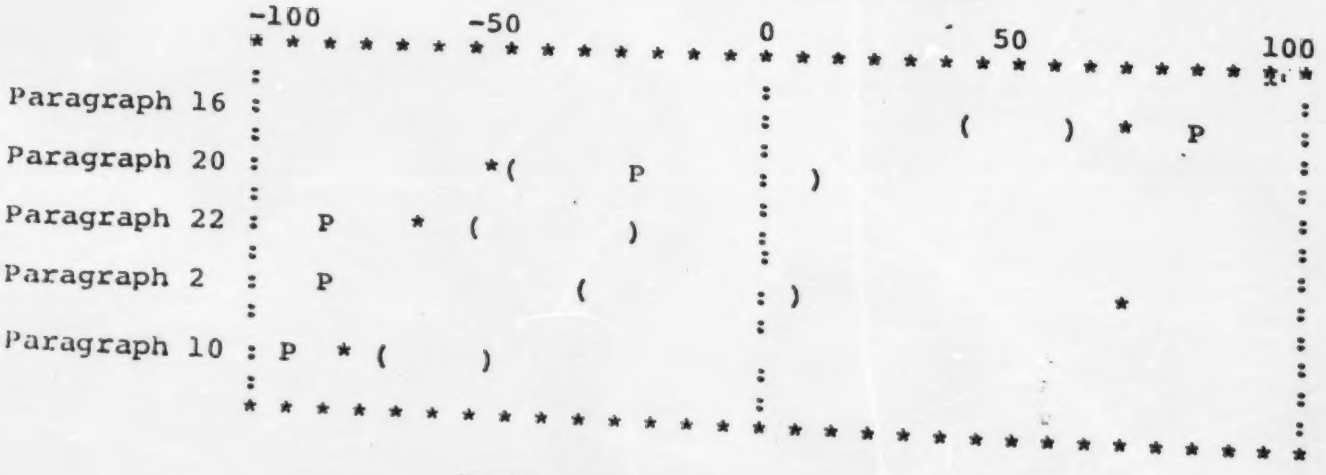
2. This technician often made personal phone calls, some of them lasting for extended periods.

10. This technician continually roamed around the building striking up conversations with anyone willing to listen on subjects such as fishing, footbal, etc., disrupting the work of others.

16. This technician worked steadily, without interruptions, until he completely finished each assignment :: when finished, often asked for or found new work to do.

20. This technician sometimes took a few extra minutes on coffee and lunch breaks ;: sometimes played chess or cards a little too long.

22. This technician often read newspapers and magazines, sat watching the clock or staring out the window, when there was work to be done.



FOR THE DIMENSION WORKS STEADILY VS. WASTES TIME YOU WERE IN THE MIDDLE THIRD OF YOUR WORK GROUP
Specifically, you did perform the positive behavior in 16
You did a good job of avoiding the negative behaviors in 20 22 10
To further your improvement, avoid the negative behaviors in 2

graphs the technician's score on the dimension, scored in the positive direction. The brackets around the star show where the technician stands relative to the middle third of his peers. Being within the brackets indicates average relative performance. The third part of Figure 2 graphs the technician's paragraph relationships. Paragraphs are ordered by dimension and by behavior effectiveness within dimension. A positive score means that the technician's behavior was rated as similar to the behavior in the paragraph. The capital P indicates how effective the behavior in the paragraph is. Thus, it is as good to be positively related to paragraphs containing effective behaviors as it is to be negatively related to paragraphs containing ineffective behaviors.

As with the standard and BAS forms, the BOS dimension score indicates average performance. However, a look at the paragraph scores shows why this technician was rated average in spite of being rated as similar to the good behaviors and the opposite of most negative behaviors. He was rated as similar to the paragraph concerning personal phone calls where other peers were far off on the negative side.

In short, BOS provide additional information concerning specific worker problems and accomplishments that is simply not available with adjective anchored or BAS. Although a relatively recent invention, three kinds of BOS have been

described in the literature. The factor type (Latham and Wexley, 1977) uses factor analysis of supervisory ratings of workers on job elements to derive the scales and requires many employees (minimum 200) to derive the scales. The item type (Latham, Fay, and Saari, 1979) uses a form of item analysis, cutting the worker requirements down some, but a minimum of about 60 is still needed. A third type of BOS (Janz, 1980), uses a factor analysis of subjective estimates of the relationships among job behaviors to derive the scales. It is called Behavior Relationship Scales (BRS) and can be completed for the ultimately small job size of a single incumbent managed by a single supervisor. When compared with BAS and adjective-anchored scales, (Janz, 1980), BRS demonstrated small improvements over adjective scales and about the same ratings characteristics as BAS. The BRS type of BOS are not cheaper than the factor or the item types, but it can be used when the number of job incumbents is small.

The costs of producing BAS and BOS are about the same and they are equally effective in terms of the numbers they produce for making personnel decisions. The key difference is that BOS have been demonstrated to do a better job of pointing up specific performance accomplishments and shortfalls. Also, since BOS focus on how often workers perform in specific ways, the process for attaching numbers

to workers is clear and BOS offer the firmest basis for mutual bargining on where a worker stands. If skillfully done, this bargining can lead to high acceptance instead of suspicion about the recorded data.

Performance Data Types: How to Decide

If the sole reason for gathering performance data is to make personnel decisions, the only objective for a performance data system would be to produce numbers that accurately reflect, and are seen to accurately reflect, worker effectiveness for the least cost. End result measures, if uncontaminated, do that best. Unfortunately, sources of contamination are pervasive and a careful look at end result measures is necessary. If sources of contamination can be measured, they can sometimes be statistically controlled.

In any case, if an integrated performance data and development system is desired, or if scorecard measures free of contamination cannot be found, BOS represent the performance data option having the best return on investment. They are flexible enough to provide accurate numerical data for making personnel decisions while the basic building blocks of BOS, the behavioral job elements (or paragraphs), serve well as an outline for a constructive performance development discussion.

Theoretically, there were two ways to accomplish

personnel development-- training and performance appraisal. It is time to take a long hard look at performance appraisal.

To appraise, according to Funk and Wagnall's (1976), is "to make an official valuation of." Performance appraisal is intended to be the process of communicating this "official valuation" to workers. The hope is that workers will then want to improve in areas they are rated as performing poorly. Unfortunatley, the very act of official valuation runs directly counter to worker acceptance, and the ultimate goal of reducing ineffective and promoting effective behavior is lost. In a series of classic studies done at General Electric, Meyer, Kay and French (1965) found that criticism produced resentment and resistance to improvement while positive feedback had no effect either way. They noted that performance topics criticised in a typical appraisal session showed even less improvement after 12 weeks than performance topics not discussed at all. As long as workers react with anger and resentment at being valued, focusing their attention on who has what numbers, performance appraisal has no practical value. It is the very essence of performance appraisal that defeats the goal of personnel development. Lets get rid of performance appraisal.

Instead, Meyer, Kay and French (1965) point to the

importance of setting specific performance goals. The participative adoption of specific performance topics, without finger waving or even showing a worker how he compares with others, was found to result in specific performance improvements. What Meyer et al. omitted was a system for insuring that relevant, specific behavioral topics be the basis for performance goal setting. They left it up to the supervisor and the worker to on-the-spot define what aspects of the job were most in need of goal setting. This lack of structure leaves supervisors having different value systems a lot of room for interpretation. By ruling out any kind of performance data, Meyer et al (1965) leaned too far towards the mayhem documented for the unstructured selection interview. A solid skeleton for a constructive performance discussion is needed. The behavior job paragraphs of the BRS type of BOS are seen as ideal for the structuring task. They describe in the workers' own terms the basic behavioral underpinnings of the job, removing the ambiguity of trait defined-adjective anchored scales. They insure that each of these behavioral topics is considered for each worker, removing the hit and miss of blank page or report card data. They avoid the contaimination often suffered by end result measures. Finally, the behavior paragraphs avoid the problems of behavior recency and scope suffered by behaviorally anchored scales.

The following section outlines a BOS based performance data and development system that faces up to the challenges of producing: a) clear and accurate performance data that is respected instead of ridiculed by workers, and b) a flexible but firm outline for a performance development discussion that results in maximum behavioral improvement.

The Performance Data and Development System

The core structure of the system is a set of BOS. Of the three types of BOS mentioned earlier--factor, item and BRS, BRS are chosen since: a) they can be operationalized for jobs having many or few incumbents, and b) the basic unit of analysis is the behavior parahraph, about 20-30 per job, vs. the behavior job element, about 60-90 per job. The smaller number of basic units for BRS make the negotiations between workers and supervisors over performance data or goal attainment a more reasonable task. Consistent with the development sc far, the system is broken into two subsystems: the performance data subsystem and the performance development subsystem. While BRS serve as the core structure for both, there are fundamental differences in how the BRS are used that take into account the conflicting goals of performance data and development. The clearer, the more specific, and the more accurate performance data becomes, the more it interferes with performance development. Thus, while ratings on BRS produce

high quality performance data that can be constructively discussed to resolve superior-worker conflicts, ratings on BRS would only serve to snarl up the performance development subsystem. For performance development, only a focus for discussion is needed, not ratings. Why cling to numbers that are not needed anyway?

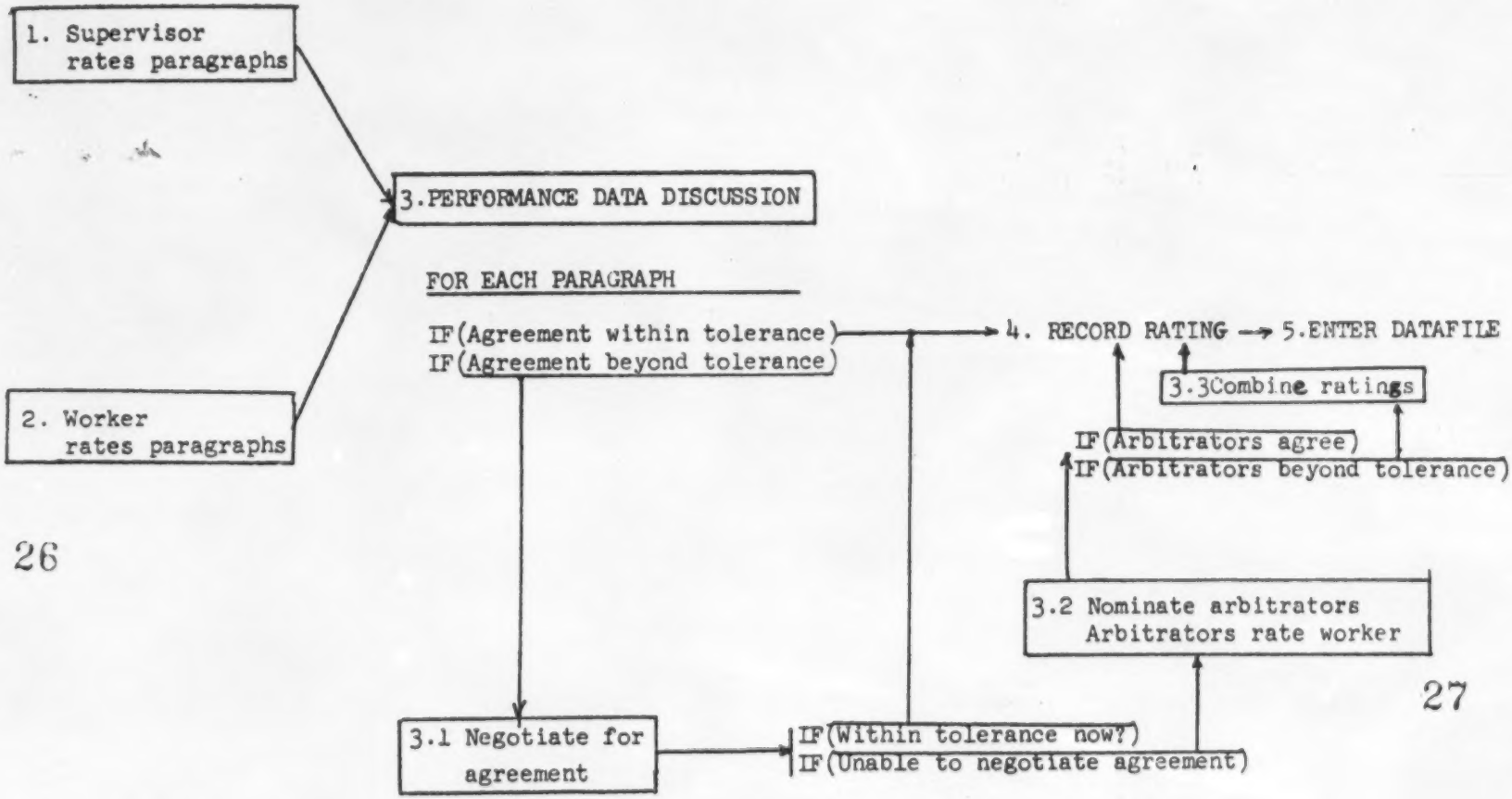
Once BRS have been developed following the steps outlined in Janz (1980), the performance data and development subsystems can be put into operation following the steps outlined in Figures 3 and 5.

The Performance Data Subsystem

Insert Figure 3 about here

Once the supervisor has rated the worker and the worker has completed self ratings, they meet for a performance data discussion. In this discussion, the worker's self rating and the supervisor's rating are treated as equally valid views of the true frequency of the behavior in question. Figure 4 illustrates some behavior paragraphs for detail design technicians and the kinds of ratings supervisors and workers would be asked to make.

Figure 3: The Performance Data Subsystem



Insert Figure 4 about here

On paragrpahs where both supervisor and worker ratings fall close enough to be within a preset tolerance level, the average of the two ratings is entered and need not be discussed. On paragraphs where the ratings fall far apart, the supervisor and worker first attempt to close the gap by referring to their recollections of the specific behavior detailed in the paragraph. If after a reasonable debate they still cannot move to within the tolerance limit, the supervisor and the worker each nominate an alternate who they feel is qualified to rate the worker's behavior on the paragraph in question. Later, if the nominees agree on the rating, their average is entered in the file. If the nominees also disagree, the average of all four ratings is entered as a last resort. Hopefully, things will not often go this far. How far things do go will depend a lot on how skillful the supervisor and worker are at resolving rating conflicts.

The extra effort towards negotiation and conflict resolution may at first blush seem to be an unwarranted expense, but keep in mind the payoffs. First, by treating worker input with actual as opposed to verbal respect, it reinforces worker feelings of self esteem, responsibility,

JOB: Detail Design Technician Name:

1. This technician put too many items on a plan sheet: had to redraw pictures of a typical section, even though he was given an example to follow: did not do a complete finished job on coloring a layout or plan: drew detail of a culvert and roadway profile so light that it could not be reproduced: erased and relocated items on a plan sheet: left holes in the paper when erasing mistakes: drafted plan sheets that were wrinkled, blotched, and difficult to read.

   Out of 100 plans or drawings, how many completed by this technician contained problems like those described above? _____

2. This technician often made personal phone calls, some of them lasting for extended periods.

   On the average, how many minutes a week did this technician spend on personal phone calls? _____

3. This technician used proper standards to establish sight distances for a roadway profile, to set roadway and shoulder widths, to set storm sewer grades taking into account utility installations, and to calculate sewer runoff rates.

   Out of 100 assignments, how many did this technician complete using the appropriate standards as outlined above? _____

4. This technician sat watching the clock: read newspapers and magazines: satstaring out the window when there was work to do.

   On the average, how many minutes a week did this technician behave as described above? _____

5. This technician worked steadily until he finished each assignment: asked for or found new work when he finished his work assignment: willingly accepted voluntary assignments.

   Out of 100 assignment, on how many did this technician behave as described above? _____

and organizational commitment. Also, respecting worker input reduces suspicion and resentment about personnel files and personnel decisions. Second, by taking two independent and sometimes conflicting views of one person's performance, the final ratings will more accurately reflect actual performance. Again, the importance of maximizing genuine worker acceptance of performance data cannot be understated if important personnel decisions such as compensation and merit pay are based on the data.

The Performance Development Subsystem

The performance development subsystem represents a systemmatic, planned approach to the setting of specific performance goals. It is a form of management by objectives, but instead of setting targets on end result measures, the focus of goal setting is on job behavior. As in the performance data subsystem, facing up to the naturally occuring conflict between worker and organizational goals is a part of the system. Performance development discussions should occur as needed, but probably should occur more often than the annual performance data discussion.

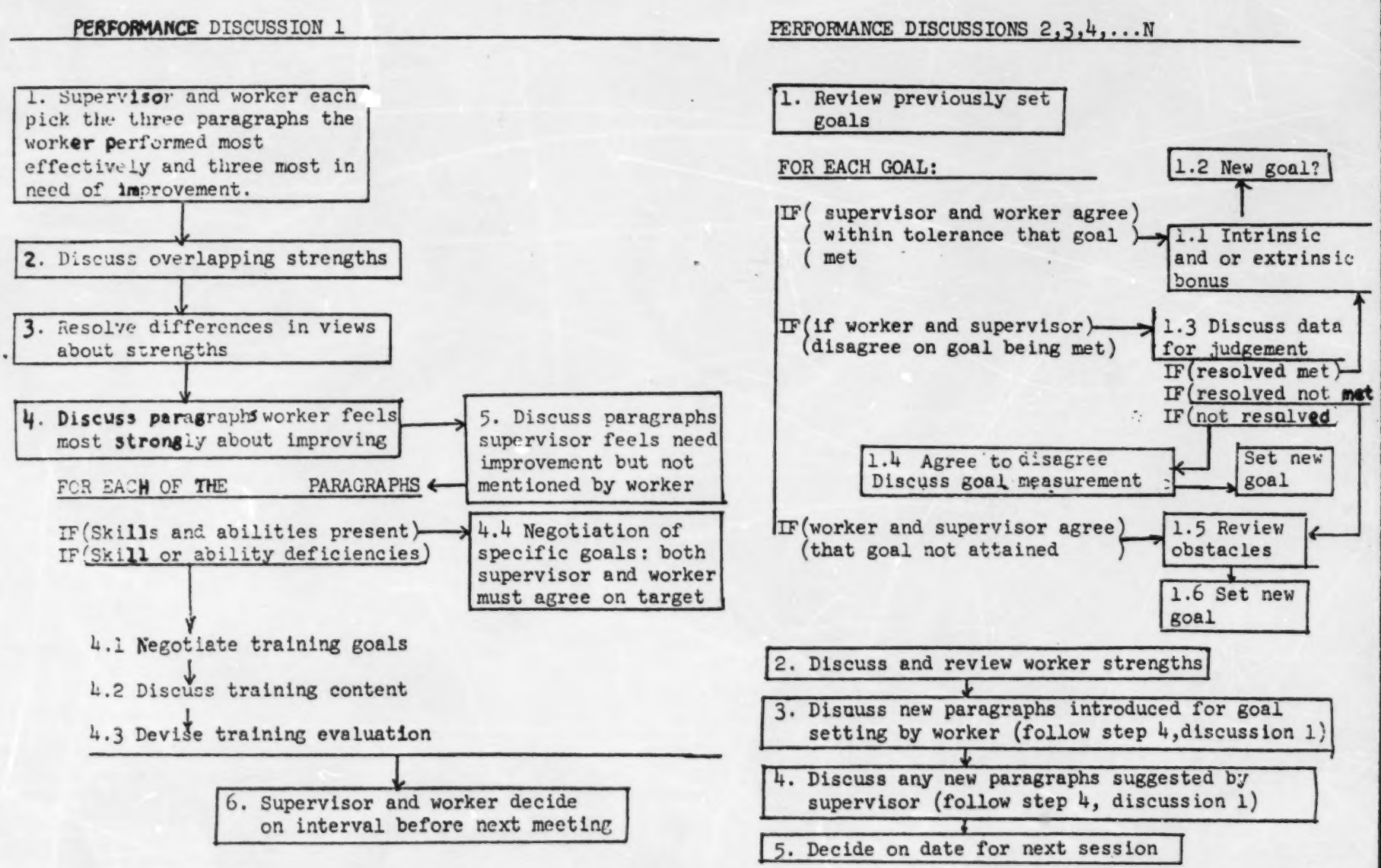
Figure 5 divides the system into two parts, one for the initial performance development discussion and one for all discussions that follow.

Insert Figure 5 about here

Preceeding the initial discussion, the worker and supervisor study the behavior paragraphs developed as a component of the BRS procedure. They are asked to pick the worker's best three paragraphs, either negative behaviors the worker avoided or positive behaviors performed frequently. Next, both the worker and supervisor independently pick three paragrpahs most in need of improvement.

To kick off the first performance development discussion, the worker's self perceived strengths are discussed along with the strengths perceived by the supervisor. The objective of this step is a healthy airing of different views, leading to a discussion and resolution of differences in those views. If a worker sees a strength not perceived by the supervisor, the worker is expected to cite some incidents when he demonstrated the strength. Thus the supervisor gains a more complete view of the worker's behavior. If the supervisor sees a strength missed by the worker, the supervisor cites examples of when that strength appeared. The discussion of strengths serves as a warm up for conflicts over where improvements are needed. It is much easier to build a good atmosphere for and skill with conflict resolution when debating strengths.

Figure 5 : The Performance Development Subsystem


PERFORMANCE DISCUSSION 1
1. Supervisor and worker each pick the three paragraphs the worker performed most effectively and three most in need of improvement.
2. Discuss overlapping strengths
3. Resolve differences in views about strengths
4. Discuss paragraphs worker feels most strongly about improving
5. Discuss paragraphs supervisor feels need improvement but not mentioned by worker
FOR EACH OF THE PARAGRAPHS
IF (Skills and abilities present)
IF (Skill or ability deficiencies)
4.4 Negotiation of specific goals: both supervisor and worker must agree on target
4.1 Negotiate training goals
4.2 Discuss training content
4.3 Devise training evaluation
6. Supervisor and worker decide on interval before next meeting
PERFORMANCE DISCUSSIONS 2,3,4,...N
1. Review previously set goals
FOR EACH GOAL:
1.2 New goal?
IF (supervisor and worker agree) (within tolerance that goal) (met
1.1 Intrinsic and or extrinsic bonus
IF (if worker and supervisor) (disagree on goal being met)
1.3 Discuss data for judgement
IF (resolved met)
IF (resolved not met
IF (not resolved
1.4 Agree to disagree Discuss goal measurement
Set new goal
IF (worker and supervisor agree) (that goal not attained)
1.5 Review obstacles
1.6 Set new goal
2. Discuss and review worker strengths
3. Disuuss new paragraphs introduced for goal setting by worker (follow step 4,discussion 1)
4. Discuss any new paragraphs suggested by supervisor (follow step 4, discussion 1)
5. Decide on date for next session

Once differences over strengths are discussed and resolved, the next phase is a discussion of the three paragraphs the worker feels most like improving. The discussion starts with the worker suggesting a specific, behavioral target to shoot for. The supervisor then agrees, or bargins for a higher target. Both worker and supervisor must agree to a specific goal before it is entered onto the performance development report.

After setting goals for the three paragraphs raised by the worker, the supervisor addresses paragraphs seen as needing improvement but not already mentioned by the worker. He raises the issue descriptively, not evaluatively, by saying: "One paragraph I would like to discuss is paragraph X. What do you think a reasonable goal on paragraph X is for you?"

Finally, supervisor and worker discuss when the next session should take place. The interval should allow enough time for the results of worker effort towards meeting the goals to take hold, but not so long that the goals become stale. This judgement is best left to the worker and supervisor instead of fixing an automatic rating period.

Notice that at no time during this initial session are workers appraised. Neither is their performance compared to other workers or criticised in any way. Thus, the resentment hostility, and resistence to improvement that is

an essential partner to performance appraisal is avoided. Instead, the session focuses on producing worker commitment to specific performance improvements.

For the subsequent performance discussions charted in the right half of Figure 5, the process begins with a discussion of goals set at the previous session. Since the supervisor and worker may disagree over whether the goal was met, the system includes a mechanism for facing up to that potential conflict.

If both supervisor and worker agree the goal was met, the accomplishment is noted and whatever rewards the organization may want to attach to goal attainment are awarded. This paper avoids speculating on specific reward contents or mechanisms. If both supervisor and worker agree the goal was not met, they discuss possible obstacles to goal atainment and re-negotiate the goal. If supervisor and worker disagree on goal performance, they discuss specific examples to support their respective views. If after a reasonable debate they still cannot agree they agree to disagree acknowledging that reasonable people can fail to reach agreement. They then discuss ways to sharpen their goal measuring behavior so that they will have an improved chance at agreement the next time they meet and discuss this specific paragraph. Further, they discuss whether they should stick with the goal that was set for this session or

mcdify it in some way.

Once the discussion of goals set in the past is completed,the supervisor and worker discuss paragraphs they feel represent strong performance. Then the worker is asked if he wants to discuss any new paragraph for goal setting. Next the supervisor may suggest a new paragraph for discussion and goal setting. Last, they discuss when the next session should be held. If some paragraphs are not agreed on, they are carried over to the next session.

Summing up, a performance data and development system that separates the conflicting objectives of personnel decision making and personnel development was outlined. The system met head on the challenge of obtaining worker acceptance both to the performance data and performance development components of the system. The system makes a lot of demands. First, it requires a careful, behavior based, job analysis be completed using the BRS system. Second, it calls for records keeping tc coordinate the meetings and the data that comes from them. Fortunately, even small ccmputers can perform most of the organizing, retrieval, and report generating functions. Third, the system calls for the development and practice of effective communication and conflict resolution skills. In their absence, the system could not be sustained.

Those are the costs. On the benefit side the system

produces clear, accurate, and relevant performance data for personnel decision making that will be accepted by workers and supervisors alike. Second, the system promotes the self esteem, interpersonal competance, and organizational ccmmitment of all those using it. Hard cast impacts of this benefit include reduced absenteeism, turnover, increased willingness to accept overtime, willingness to accept unpleasant tasks, willingness to handle peak loads, willingness to bargin in good faith, and other specific behaviors too numerous to detail. How long will management go on paying for hanging on to performance appraisal systems that promote resentment instead of commitment, bickering instead of problem solving, and resistence instead of mutual agreement on specific performance improvements? As long as managers value simplicity over facing up to complexity, the use of power over the use of reason, and numbers over people. The money is waiting to be saved but managers must be willing to take the risks.

References

Bernardin, H.J. Behavior expectation scales vs. summated scales: A fairer comparison. Journal of Applied Psychology, 1977, 62,422-428.

Borman, W.C. and Dunnette, M.D. Behavior based vs. trait oriented performance ratings: An empirical study. Journal of Applied Psychology, 1975, 60, 561-565.

Janz,J.T. Behavior relationship scales: Behavior observation scales for small samples. Simon Fraser Discussion Series 80-12, Department of Business Administration, Simon Fraser University, 1980.

Latham, G.P. and Wexley, K.N. Behavior observation scales. Personnel Psychology, 1977, 30, 255-268.

Latham, G.P., Fay, C.H., and Saari, L.M. The development of behavior observation scales for appraising the performance of foremen. Personnel Psychology, 1979, 32, 299-311.

Meyer,H.H., Kay, E. and French, J.R.P. Split roles in performance appraisal. Harvard Business Review, Jan-Feb 1965,p123.

Smith, P.C. and Kendall, L.M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 1963, 47, 149-155.